DOCUMENT RESUME

ED 212 658 TM 820 052

| | |
|---|---|
| AUTHOR | Baker, Eva L. |
| TITLE | Recommendations for Training of Teachers, Parents, and Other Constituencies in the Use of Tests. Studies in Measurement & Methodology, Work Unit 1: Design and Use of Tests. |
| INSTITUTION | California Univ., Los Angeles. Center for the Study of Evaluation. |
| SPONS AGENCY | National Inst. of Education (DHEW), Washington, D.C. |
| PUB DATE | Nov 79 |
| GRANT | OB-NIE-G-78-0213 |
| NOTE | 21p. |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Achievement Tests; Educational Improvement; Elementary Secondary Education; Examiners; *Teacher Education; *Testing; *Training |

ABSTRACT

The general topic of training needs related to achievement testing is addressed. Questions are raised about training as a means for educational improvement; needs specific to the achievement testing area are discussed; and a specific list of questions to be considered in planning training efforts is presented. It is concluded that using a thematic orientation, perhaps of communication, instruction and testing practices might be reworked so that what happens to students in classrooms occurs as a natural process rather than a series of abrupt and disjoint enterprises. Similarly, it is recommended that training audiences be integrated, so that all participants can understand the roles of one another and can formulate reasonable expectations for team performance. Such integrating of practices would mitigate against isolated "workshop" type experiences for insular audiences. The challenge is to develop or to share already existing successful training tactics, and to fuse them into a sensible and continuing program for improving the effectiveness of schools. (Author/GK).

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED212658

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE: NOVEMBER 1979

STUDIES IN MEASUREMENT & METHODOLOGY

Work Unit 1:
Design and Use of Tests

Eva L. Baker & Edys Quellmalz
Project Directors

# RECOMMENDATIONS FOR TRAINING OF TEACHERS, PARENTS, AND OTHER CONSTITUENCIES IN THE USE OF TESTS

Eva L. Baker

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.S. Gray
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Grant OB-NIE-G-78-0213

TM 820 052

Center for the Study of Evaluation
Graduate School of Education
University of California - Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

This report addresses the general topic of training needs related to achievement testing. It begins by raising questions about training as a means for educational improvement, discusses needs specific to the achievement testing area, and finally presents a specific list of questions to be considered in the planning training efforts.

Technical assistance and training permeate social and educational services. When programs show little success, policy makers, borrowing from the formerly more prestigious area of foreign aid, often impose training requirements to shore up spotty program use. While no self-respecting social analyst would publically admit the belief, the adage that "knowledge is power" underlies the training alternative, that if only constituencies understood the innovative ideas (or better still, the concepts), new programs would more easily be adopted and supported.

Unbridled optimism aside, why is training a preferred course of action for program improvement? For one point, we opt for training when the alternative is costly and painful program redesign. It is infinitely simpler to develop adjunct training programs than to rethink from top to bottom, confused or unsuccessful programs. Supported by cost considerations and the psychic involvement of program creators, inertia or patching-up tactics such as training sometimes conspire to lengthen the lives of programs more properly revised or discarded. Training apparently repeats a pattern of many social services: blaming the victim. The program user bears the responsibility for program failure, and through lack of program information, skill or motivation is thought to inhibit the progress imagined by program developers.

The relationship of training to the area of achievement testing reflects this general orientation, particularly because of the growing prominence of testing itself in the lives of both school

personnel and bureaucrats. Testing is on the upswing. Previously, training in educational measurement used to be limited to those individuals who planned lives of research (and presumably reflection) and who, by either personal quirk or unflagging diligence, managed to keep themselves interested in this arcane topic. As the Federal establishment, shadowed by State efforts, tied more and more funding to evaluation activities, educators needed more expertise in the use and design of tests. Staffs of state departments of education, school districts, and particular schools were regarded as proper targets of measurement training ventures. Recently, however, concerns for training in testing have been directed to a broader audience. Because of the visibility of laws requiring student competency for graduation or grade-to-grade promotion, teachers are expected more than ever to understand tests, to make use of their findings, and to demonstrate how teaching improves because of changes inspired by formal test information. If test results, over time, do not demonstrate positive changes, teachers will be the prime suspects. Teachers are thought to lack the information, skills, and motivation needed to improve students' performance, i.e., test scores. The quality of the tests themselves, or, in fact, the quality of the idea that teachers formally account for test results in teaching do not impede the rush to test-teach-test advocacy. Only teacher organizations and some university folk have explicitly raised these questions, but the charge of self-interest vitiates their concern.

The tacit acceptance by many educators that training in testing is just what is needed, and the resultant oversimplification about the

classroom use of tests, perversely feed and are nourished by the current (and, no doubt, perennial) controversy on technical issues in testing. Contention surrounds almost every imaginable combination of issues. We argue over technique, format, language, syntax, bias, standards, administrative conditions, and legalities. Faint mutters filter through on whether the educational community, as a whole, actually profits from such tests, whether tests have the power to influence instruction in the manner promulgated, and whether tests are the best investment to improve school outcomes. Similarly gently wafting are ideas about who should use test results, what translations are desirable, what supporting structures and materials should be in place, and how such use should be scheduled. We seem to assume, for instance, that teachers should apply test results as often as possible to inform their efforts. We value frequency. And testing zealots, waving IBM answer sheets, prevail. We find school districts imprinted with their effects, school districts which pretest children on Mondays in all areas of instruction, and posttest them on Friday afternoons. (They use the weekend, between barbecue and tennis, to inspect the data, to see how effective teaching was for the one week interval, and presumably, to revise teaching plans.) What such practices do to the rhythms of instruction, the anxieties of children, and the social roles of teachers remains relatively unacknowledged and substantially unstudied.

If training teachers in the use of tests seems like a good idea, a better idea following fast is to train other members of the community as well. Not only may teachers profit from learning how to employ

tests in an optimal (but presently unknown) manner, but others are thought to need this information as well. Knowledge of tests should be shared, it seems, with parents, community members, school administrators cloistered too long in plots over declining enrollment, and school boards. Certainly we could also give a subtle education to the media, especially to those newspaper reporters who continually embarrass the educational establishment with their periodic ranking and publication of schools' test scores. We should also attend to legislators and their staffs (those with the most power and frequently the least information about how tests work and what may be expected of them). Notice that once a training alternative is adopted, the audiences to which we generalize expand$.

The training of teachers in test use probably cannot be avoided. Legal precedent requires appropriate notice of school system personnel (and their client students) when new testing requirements are imposed, and teachers need to be informed about content and purposes of tests. School districts, to behave in an acceptably accountable manner, appear to support these training requirements. Data generated by research studies corroborate that teachers do not have much information about tests, and by and large, do not incorporate test results in their teaching plans. But before launching into a discussion of how and what should be trained, we must take account of three important cautions:

1. Most available tests have not been developed in a way that allows teachers to make clear inferences for instructional action.

2. Consequently, almost no hard evidence exists that training teachers in test use improves instruction.

3. Perhaps teachers have good reason for their disinterest in and low use of test scores.

Suspending these cautions for the time, this paper will address the procedures one might wish to use to go about training educational constituencies, even though research and reflection might lead us in the future to second-guess the wisdom of the testing enterprise. Despite social strides made in sensitivity to the use of deficit models to explain behavior of the culturally different, the establishment nonetheless imposes such a model to justify training teachers in testing. In-service* recipients, teachers, may be so disheartened that they don't even notice the slur. A basic set of questions guides our approach to training.

1. Who is sponsoring the training? For what explicit purposes? For what implicit purposes?

2. Who is to be trained? How serious are the training goals taken? What motive do the trainees have for participating?

3. What is to be "trained"? Is the training to impart information for general use, to prepare individuals to exhibit skills, to modify general attitudes and predispositions?

4. What means are selected for use? How likely are the means to accomplish the goals of training? How will one know if training is successful?

5. What supports are needed or available in the trainees' regular setting, e.g., classroom, district office, to enhance the training effort?

6. What alternatives are there to training as the means to improve educational practices?

---

*A peculiarly bovine term.

Let's explore some ways in which these questions might be answered. (They are not mutually exclusive as the brief discussion below will demonstrate.)

Who is sponsoring the training?

Is a federal agency supporting this work? They may do so out of the belief that training itself will result in improved practice; they may support such efforts because of need to demonstrate capacity to respond to school district concerns. At the school district level, training in testing may serve as surrogate for other, more systemic actions. It may be a cheaper alternative than new programs. On the other hand, training may augment a curriculum which depends upon iterative testing cycles. Training may be sponsored for even more bureaucratic reasons, for instance, to meet expectations about the district's role in staff development or to provide a more "accountable" image for the public. Statements of acknowledged expectations should be created and should extend beyond a simple recitation of desired skills. If teachers are expected, over time, to change habitual classroom practices, such objectives should be made clear.

Who are to be the recipients of training?

If teachers are the principal audience of training, reasons for the experience should be formulated. Training may be selected because particular difficulties have been experienced in given arenas of instruction: improving the performance of poor children, implementing a new curriculum, relating and justifying students' grades to the parent community, or perhaps responding to newly legislative requirements for competency testing. Has the trainee group been selected because it is especially needy, or perhaps, especially open-minded? If parents are involved in training, is the goal awareness, or more

on the order of specific activities designed to help their children learn? The incentives to take training seriously should also be explored? Benefits for acquiring desired skills demand articulation, and may in fact, be difficult to enunciate. Sanctions, if any, minimally require identification.

What is to be "trained"?

Describing how the goals were selected and any evidence relating to their validity represent minimum effort. Have expectations been stated regarding specific accomplishments? Trainers should also suggest the extent to which information presented is expected to generalize beyond the particular training setting. (If training is presented in mathematics testing, are principles taught useful in language testing as well?) Who decided on the goals?

What means are used in training? How is success assessed?

Staff development activities in educational bureaucracies routinize quickly. We need fresh approaches, employing the general principles of learning in new combinations. Clarity, structure, time on task, locally relevant problems, and enthusiasm blend with pacing, scheduling (oh, no, not after school), and chunking (size of instructional segment) to influence training success. If training involves more than one audience, what unit is best selected? One option would involve training all teachers in particular subject matter areas, allowing the presentation of problems and examples of special pertinence. In contrast, training groups might mix teachers from different grade levels and subject specialties. Training might well be approached as a school level activity with teachers, parents, administrators and interested community participating together, with whatever risks and threatened vanities such an arrangement may produce.

Format questions are related to the selection of a training leader. How credibility of leadership is treated, competency conveyed, and authority (if necessary) portrayed may be features critical to success. In addition, we need to consider how such training is evaluated, for many efforts have foundered at the outset by clumsy scientism, with an overlong and underexplained pretest, initiated in the name of evaluation. Evaluators will decide on critical trade offs in techniques balancing the precision of measurement, the reality of work samples, and the safety of self-reports.

How much time should elapse between training and evaluating its effectiveness? What dependent measures appear of most utility?

Delicacy is required in the evaluation of any training effort. Typically, no such evaluation occurs and the "success" of the endeavor is inferred when teachers do not rise up in mutiny. How we may assess the impact of particular training activities relates directly to the goals of the training, as well as other potential consequences. Thus, efforts in evaluation ought to be commensurate with the anticipated impact and resource allocation in the training effort.

What training supports are available?

For any instructional "treatment" to last, the practice period must extend beyond formal training. Follow-up materials, exercises, and in-class problems should be available and involve the application of general principles to specific problems. A system of sharing and feedback might involve pairs of trainees providing peer support and review, or might involve a more hierarchical arrangement.

Support for training can take a concrete and positive form in the kinds of materials available for teachers to use. Imagine two teachers each suitably "trained" and eager to apply tests in instruc-

tional decision-making. One teacher returns to the classroom and tries to impose the logic of test-based decision-making upon the usual environment. For a teacher to apply such principles, he/she will need test data reported on a known schedule, in a careful format, time to interpret the results, energy and skill to select from variously arranged curricula those lessons most suitable to improve students' learning,* and help in managing the entire affair, especially when students' individual differences show up in results, as they are most likely to do.

On the other hand, another teacher might have much greater success in test information use if the connections between test and instruction are already made. Certain curricula include tests presumably closely related to the planned curricula. Although on a technical level many of these tests could stand improvement, they relieve the teacher from the obligation of searching and matching instruction to test results. The provision of "matched" instruction to tests has taken many forms. Systems are in practice that attempt to key extant curriculum materials and texts to particular tests. Other test developers have created their own set of instructional resources and practice exercises for teachers. Rather than commitment to a given test, such systems build on teachers' preferences for particular materials and weave test use into established curricular choices. But whether the point of entry is creating a test and finding curricula in the commercial sector, or actually developing matched instruction, the teachers' job is simplified. Note, too, that the attempts to find extant tests and match them to extant instruction seems to falter under intensive research analysis. The matches are just not there.

---

*Presently only a weak-knowledge base is available.

At what point should teachers be involved?

Analysts of change often emphasize the importance of "ownership" or "buying in" to the process. The level of support provided for teachers, if real differences between future and present practice is desired, may require involvement by teachers at very early and continuing times. The test as a product of great wisdom and dazzling technology probably cannot simply be presented to and adopted by teachers. Devising a marketing approach to test use is an alternative, but one which may worry us on an ethical level. To build commitment, we could have teachers participate seriously, not ceremonially, in the design of tests and instruction for use in their schools. Painful as this process will be and costly as we may fear, experts believe that skills and beliefs grow with sustained involvement, when teachers and others have some responsibility for creating what they use. Clearly, the burden that training bears may be greater than either we imagined or for which we have funds.

What are the alternatives to training in test use?

Alternatives can be explored only with knowledge of constraints to their use. If we assume limited teacher time and good will, scarce resources, and pressure to show noticeable results, general training in measurement ideas does not seem to be a likely winner. The amount of effort and dollars might better be spent in a different way. Suppose, for instance, that a school is located in an area which is subject to great mobility. Perhaps the money spent on training teachers in testing might be better directed to helping teachers explore a range of

Goals for what purposes?

Usual discussions of goal selection seem to depend upon a formal process, such as needs assessment, where potential participants are asked their preferences and given the option to influence programs in which they will have a stake. In technical areas, however, we should question the utility of needs assessment tactics. The level of understanding of the participants clearly limits the extent to which they are aware of the information needed to make good use of tests. They may have developed general impressions, ready reactions, and some successful procedures, but by and large, most groups, teachers included, have not had much technical background related to testing. One way in which to approach this problem is to subdivide goals into categories which at once direct the means of training to a greater degree and at the same time help one select best audiences for given intents.

One scheme which can be used is to look at test information and determine who is to benefit by its use. For instance, we can say that some test use directly affects the data providers, the students who took the test. This level one use (Baker, 1979) is illustrated by competency and grade-to-grade promotion tests as well as some placement and diagnostic tests. For such tests, parents and community members need general awareness of the likely consequences of such tests. Teachers will also need relatively intense training to enable them to prepare instruction that provides a reasonable opportunity for students to succeed. Administrators will also have to understand what resources

may be required for the implementation of a rather different instructional program. The media could also use information about what inferences should be made about test performance.

CHART I
Training About Tests that Affect Student Examinees

| Audience | Classification of Goals | Length of Training Program |
|---|---|---|
| Students (as maturity permits) | .. Awareness<br>.. Study Skills | Brief<br>Concurrent with subject matter instruction |
| Teachers | .. Skill Development:<br>-planning instruction<br>-interpreting test data<br>-sharing resources | Intensive and continuing |
| Administrators (Counselors) | .. Awareness<br>.. Management Options | Intermediate |
| Parents | .. Awareness<br>.. Adjunct Instructional Skills for Support | Brief<br>Intermediate |
| Community (Governance Structures) | .. Awareness | Brief |
| Media | .. Awareness | Brief |

A second level of test is one whose use is intended to influence instruction in general. Decisions regarding presentation, pacing and materials for instruction may be inferred from test performance, sometimes collected in a program evaluation framework. In this context, the principal participants in training ought to be those school personnel directly responsible for teaching and providing resources for instructors.

CHART II
Training About Tests that Affect Instructional Planning

| Audience | Classification of Goals | Length of Training Program |
|---|---|---|
| Teachers | .. Instructional planning<br>.. Materials selection | Intensive and Continuing |
| Administrators | .. Data interpretation<br>.. Management Options | Intensive |
| Parents | --- | --- |
| Community (governance structures) | .. Awareness | Brief |
| Media | .. Awareness | Brief |

A last or third level of test use involves serious policy inferences derived from data about program operation, general pressures from the community, or expectations from the legislature. In these decisions, teachers are much less directly involved, although the consequences of policy choices inevitably affect teachers dramatically.

CHART III
Training About Tests that Affect Policy Decisions

| Audience | Classification of Goals | Length of Training Program |
|---|---|---|
| Teachers | .. Awareness | Brief |
| Administrators | .. Data interpretation<br>.. Technical support | Intensive |
| Community (governance structures) | .. Data interpretation<br>.. Technical support | Intermediate |
| Parents | --- | --- |
| Students | --- | --- |
| Media | .. Awareness | Brief |

Certainly rough charts, such as presented above, only begin to organize the range of alternatives. Terms such as brief, intermediate, etc., need explication. But it should be clear that (1) training is not thought to be a commodity administered in equal dosage to all audiences; (2) the broad purposes intended for test administration suggest alternative primary audiences.

From goal to objectives

For illustrative purposes, a sample of objectives derived from these general goal areas will be provided.

Teacher Skill Development (Chart I):

Objective: To identify from given test specifications instructional materials/content appropriate for student performance.

Sub-objective: To identify (and create) practice materials for the content and behavior domains presented in the test specifications.

Objective: From given sets of data, to group students who are in need of specific skill development and identify instructional sequences likely to succeed for each group.

Even relatively specific objectives such as these can require a formidable expenditure of energy and good will from both trainer and trainees. Here are additional examples:

Administrators-Data Interpretation (Chart II)

Objective: To infer from data, patterns of performance which are likely to be a) population related; b) school related; c) program related; d) teacher/classroom related.

Media-Awareness (Chart III)

Objective: To interpret data so that the choices in management options are limited, and to identify the benefits and costs (with help) associated with these options.

Clearly, the full range of reasonable objectives inferred from these general goals may be stated and then subjected to the scrutiny of those with most interest and need for participation. The selection of goals for training is similar to other curriculum problems, and the extent to which purposes for training are seriously held and sufficiently supported both administratively and economically will influence the detail and breadth of the identification of objectives.

Formats

Although briefly mentioned in the discussion of training means, the format of training is a problem that has general roots in all in-service training efforts. For example, the organizational structure selected for training will certainly influence its success: individual training (self-instructional materials) assumes that peer support is either unnecessary or easily developed. Training conducted under external-to-district auspices, e.g., university extension program, certification courses, or regular graduate programs, suggests that external rewards, salary credits linked to course experiences, for instance, are essential ingredients for success. The omnipresent "workshop" format implies that at least some short-term artifacts or immediate applications will occur as a consequence of this sort of

"craft" session. We should not attempt to imagine that teachers and others' expectations of what will occur is unrelated to the methods we promulgate for training. Similarly, the authority, credibility and experience of the trainer may require very different persons for training for different audiences.

Conclusions

These comments regarding the training of personnel for the application of test information must be again tempered by the concern that money might be better spent on the development of instruction-testing cycles which do not artificially separate the functions of teaching and assessment, but rather take express pains to link them. Using a thematic orientation, perhaps of communication, instruction and testing practices might be reworked so that what happens to students in classrooms occurs as a natural process rather than a series of abrupt and disjoint enterprises. Similarly, to the extent possible, we recommend that training audiences be integrated, so that all participants can understand the roles of one another and can formulate reasonable expectations for team performance. Such integrating of practices would mitigate against isolated "workshop" type experiences for insular audiences. The challenge will be to develop, or to share already existing successful training tactics and to fuse them into a sensible and continuing program for improving the effectiveness of schools.

## REFERENCES

Baker, Eva L. Achievement tests in urban schools: New numbers. Vol. 4. Paper prepared for the National Conference on Urban Education, St. Louis, Missouri, July 10-14, 1978.

This report addresses the general topic of training needs related to achievement testing. It begins by raising questions about ~~the~~ training ~~option~~ ~~itslef~~ as a means ~~of~~ for educational improvment, discusses ~~training~~ needs such concerns specific to the achievement testing area, and presents a specific list of questions to be considered in the planning ~~of such~~ training ~~ventures~~ efforts.